MEANS AND VARIANCES IN STOCHASTIC MULTISTAGE CANCER MODELS

AIDAN SUDBURY,* *Monash University*

Abstract

A widely used model of carcinogenesis assumes that cells must go through a process of acquiring several mutations before they become cancerous. This implies that at any time there will be several populations of cells at different stages of mutation. In this paper we give exact expressions for the expectations and variances of the number of cells in each stage of such a stochastic multistage cancer model .

Keywords: Multistage cancer model; mutation chain

2010 Mathematics Subject Classification: Primary 60J20; 92B05

1. Introduction

One model of cancer development that has been widely studied is what is generally called a stochastic multistage model (see, for example, Zheng (2008)). At a fixed rate a healthy cell, when it divides, may give rise to a healthy cell and a mutant cell. A mutant cell, when it divides, may then give birth to two mutant cells of the same type, or possibly to one of the same type and to one with a further mutation. This process may continue so that the population consists of several subpopulations of cells with each subpopulation having a different number of mutations. We represent such a model in Figure 1, where cell death has also been allowed for. The classical treatment of the two-stage model is given in Armitage and Doll (1957). However, it is now generally considered that the number of stages through which the cells have to pass before becoming carcinogenic is greater than two. Such models have been analysed using computer simulation in Connolly and Kimbell (1994) and theoretically in Portier *et al.* (1996), (2000). Previous derivations have been somewhat difficult to follow due to the mathematical complexity of the treatment. In this paper we consider a slightly simplified version of this model, allowing us to calculate exactly expectations and variances for the number of cells in each stage of the multistage model. Since it is believed that a certain number of mutations may be necessary before a tumour becomes cancerous, we shall also give a recursive formula for the probability that at time t the number of mutations has reached a certain threshold.

The major drawback of this treatment is the assumption that all cells act independently of all others, an assumption that may be reasonable while the numbers of cancerous cells are small. Nor does the treatment take into account programmed cell death or the possibility that cancer rates may change at different ages.

We imagine that there are $r + 1$ subpopulations, $S_0, S_1, \ldots, S_r$, and that the number of mutations in S_i is i. Subpopulation S_0 is different from the rest. Otherwise, cells in subpopulation S_i, $i \neq 0$, divide in two ways. First, a cell is replaced by two new cells at rate μ_i so that the subpopulation increases by 1. This can be considered a pure birth process at rate μ_i.
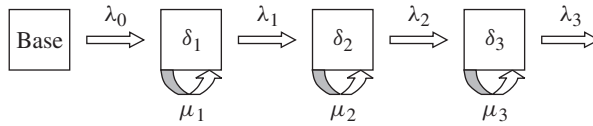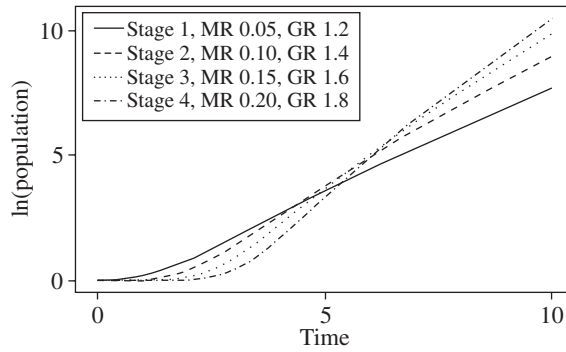
FIGURE 1.

FIGURE 2: Log population size versus time for a four-stage model with mutation rates (MRs) 0.05–0.20 and growth rates (GRs) 1.2–1.8.

Second, at rate λ_i, two new cells are produced, one in S_i and the other in S_{i+1}. In other words, each cell in S_i produces a Poisson process of cells at rate λ_i into S_{i+1}. The death rate of a cell is δ_i. The overall growth rate $\gamma_i = \mu_i - \delta_i$. All possible cell divisions and deaths are assumed independent.

The S_0 case is rather different. We assume that the subpopulations of mutants are sufficiently small compared to that of S_0, so that S_0 maintains its size, and simply produces a stream of mutants into S_1. This rate will be taken to be λ_0 which, unlike the other λs, will be for the whole of S_0 rather than a rate for an individual cell. We shall show the following result.

Theorem 1. *If $Z_r(t)$ is the number of cells in S_r at time t and $\gamma_i \neq \gamma_j$, $i \neq j$, then*

$$\mathrm{E}\{Z_r(t)\} = \lambda_0 \cdots \lambda_{r-1} \sum_{i=1}^{r} \frac{1}{\prod_{j \neq i}(\gamma_i - \gamma_j)} \frac{\mathrm{e}^{\gamma_i t} - 1}{\gamma_i},$$

and, setting $\gamma_0 = 2\gamma_r$,

$$\mathrm{var}\{Z_r(t)\} = \lambda_0 \cdots \lambda_{r-1}(\mu_r + \delta_r) \sum_{i=0}^{r} \frac{1}{\prod_{j \neq i}(\gamma_i - \gamma_j)} \frac{\mathrm{e}^{\gamma_i t} - 1}{\gamma_i}.$$

We set $\gamma_0 = 2\gamma_r$ purely to make the formula simple. It should be noted that the mutation rates $\{\lambda_i\}$ appear only in the form of their product. In Figure 2 we show the growth in the expected population sizes for a four-stage model with mutation rates into stages 1–4 of 0.05–0.20 and growth rates of 1.2–1.8. It is seen that, once established in a stage, the growth rates dominate.

2. The mean and variance of the population in S_r

We use a method employed in Kingman (1975) for the treatment of the first birth problem in branching processes. Let $B_{r1}, \ldots, B_{rm}, \ldots$ be the birth times of cells in S_r. These include

mutations from S_{r-1} and births within S_r. Let $Z_r(t)$ be the number of cells in S_r at time t, and define

$$b_r(\theta) = E\left\{\sum_m e^{-\theta B_{rm}}\right\} = \int_0^\infty e^{-\theta t} d\{E\, Z_r(t)\}, \tag{1}$$

by Fubini's theorem.

Lemma 1. *If $B_1, \ldots, B_m, \ldots$ are the birth times in S_r of a single cell in S_{r-1} born at time t_0 then*

$$E\left\{\sum_m e^{-\theta B_m}\right\} = e^{-\theta t_0} \frac{\lambda_{r-1}}{\theta - \gamma_r}.$$

Proof. In a birth-and-death process with birth rate μ and death-rate δ starting from a single cell at time 0, the expected size of the population at time t is $e^{\gamma t}$. Thus, if the number of descendants of the cell in S_{r-1} at time s is $Z(s)$ then

$$E\{Z(t)\} = \int_{t_0}^t e^{\gamma_r(t-s)} \lambda_{r-1}\, ds = e^{\gamma_r t} \frac{\lambda_{r-1}}{\gamma_r} [e^{-\gamma_r t_0} - e^{-\gamma_r t}], \qquad t > t_0,$$

and

$$E\left\{\sum_m e^{-\theta B_m}\right\} = \int_{t_0}^\infty e^{-\theta t} \lambda_{r-1} e^{-\gamma_r t_0} e^{\gamma_r t}\, dt = e^{-\theta t_0} \frac{\lambda_{r-1}}{\theta - \gamma_r}.$$

Lemma 2. *It holds that*

$$b_r(\theta) = \frac{\lambda_{r-1}}{\theta - \gamma_r} b_{r-1}(\theta).$$

Proof. Let $F_{r-1,k}(t)$ be the distribution function of the birth time of the kth cell in S_{r-1}. If this cell gives rise to cells in S_r at times $B_{rk1}, \ldots$ then, by Lemma 1,

$$E\left\{\sum_m e^{-\theta B_{rkm}}\right\} = \frac{\lambda_{r-1}}{\theta - \gamma_r} \int_0^\infty e^{-\theta t} dF_{r-1,k}(t).$$

As $b_r(\theta) = \sum_k E\{\sum_m e^{-\theta B_{rkm}}\}$ and $E\{Z_{r-1}(t)\} = \sum_k F_{r-1,k}(t)$, the result follows.

Since we are assuming that the number of cells in S_0 stays constant, λ_0 represents the total mutation rate from S_0 and we take $b_0(\theta) = 1$. Lemma 2 and a standard partial fraction expansion then imply the following result.

Lemma 3. *If $\gamma_i \neq \gamma_j$, $i \neq j$, then*

$$b_r(\theta) = \prod_{i=1}^r \frac{\lambda_{i-1}}{\theta - \gamma_i} = \lambda_0 \cdots \lambda_{r-1} \sum_{i=1}^r \frac{1}{\prod_{j \neq i}(\gamma_i - \gamma_j)} \frac{1}{\theta - \gamma_i}.$$

Using (1) and Lemma 3, we obtain the first part of Theorem 1.

If $\gamma_i \neq \gamma_j$, $i \neq j$, then

$$E\{Z_r(t)\} = \lambda_0 \cdots \lambda_{r-1} \sum_{i=1}^r \frac{1}{\prod_{j \neq i}(\gamma_i - \gamma_j)} \frac{e^{\gamma_i t} - 1}{\gamma_i}.$$

In a birth-and-death process with birth rate μ and death rate δ starting from a single cell at time 0, the variance of the size of the population at time t is $V(t) = \gamma^* e^{\gamma t}(e^{\gamma t} - 1)$, where $\gamma^* = (\mu + \delta)/(\mu - \delta)$ and $\gamma = \mu - \delta$. Thus, if $V_r(s)$ is the variance of the number of

descendants in S_r at time s of a cell in S_{r-1} born at time t_0, then

$$V_r(t) = \gamma_r^* \int_{t_0}^{t} (\mathrm{e}^{2\gamma_r(t-s)} - \mathrm{e}^{\gamma_r(t-s)}\lambda_{r-1})\,\mathrm{d}s$$

$$= \gamma_r^* \frac{\lambda_{r-1}}{\gamma_r}\left(\frac{\mathrm{e}^{2\gamma_r(t-t_0)} - 1}{2} - [\mathrm{e}^{\gamma_r(t-t_0)} - 1]\right), \qquad t > t_0.$$

Thus,

$$\int_{t_0}^{\infty} \mathrm{e}^{-\theta t}\,\mathrm{d}V_r(t) = \lambda_{r-1}\gamma_r^* \int_{t_0}^{\infty} (\mathrm{e}^{-2\gamma_r t_0}\mathrm{e}^{-(\theta-2\gamma_r)t} - \mathrm{e}^{-\gamma_r t_0}\mathrm{e}^{-(\theta-\gamma_r)t})\,\mathrm{d}t$$

$$= \lambda_{r-1}\gamma_r^*\gamma_r \frac{1}{(\theta - 2\gamma_r)(\theta - \gamma_r)}\mathrm{e}^{-\theta t_0}.$$

If the kth particle born in S_{r-1} produces a population in S_r with variance $V_{rk}(t)$ then, setting $V_r(t) = \mathrm{var}\{Z_r(t)\}$, $V_r(t) = \sum_k V_{rk}(t)$ so that

$$\int_0^{\infty} \mathrm{e}^{-\theta t}\,\mathrm{d}V_r(t) = \frac{\lambda_{r-1}\gamma_r\gamma_r^*}{(\theta - 2\gamma_r)(\theta - \gamma_r)} \sum_k \int_0^{\infty} \mathrm{e}^{-\theta t}\,\mathrm{d}F_{r-1,k}(t) = \frac{\gamma_r\gamma_r^*}{\theta - 2\gamma_r}b_r(\theta),$$

as in the proof of Lemma 2. By setting $\gamma_0 = 2\gamma_r$ we can express the formula for the variance in much the same form as for the mean. Using the above equation and Lemma 3, we obtain the second part of Theorem 1.

If $\gamma_i \neq \gamma_j$, $i \neq j$, and $\gamma_0 = 2\gamma_r$, then

$$\mathrm{var}\{Z_r(t)\} = \lambda_0 \cdots \lambda_{r-1}(\mu_r + \delta_r) \sum_{i=0}^{r} \frac{1}{\prod_{j\neq i}(\gamma_i - \gamma_j)} \frac{\mathrm{e}^{\gamma_i t} - 1}{\gamma_i},$$

since $\gamma_r\gamma_r^* = \mu_r + \delta_r$.

3. The probability S_r is occupied

In this section we will calculate the probability that the mutation chain has reached S_r by time t. In order to solve the differential equations, we have assumed that the death rate $\delta = 0$. We will start by calculating the probability that a cell born at time 0 in S_{r-1} has offspring in S_r, and then we will work backwards until we can calculate this probability for cells born in S_0. We will distinguish two ways in which descendants of the cell in S_i arrive in S_r. First, they can arrive by direct mutation from the cell in S_i to S_{i+1}. This stream of cells is a Poisson process at rate λ_i. Second, they can arrive as mutations from S_i to S_{i+1} from descendants of the original cell within S_i, that is, from the birth process in S_i of that cell. What happens from S_{i+1} is not relevant to this definition, only the manner in which descendants arrive in S_{i+1}. For $i < r$, define

- $C_i(t)$ to be the probability that a cell born in S_i at time 0 has no descendants in S_r at time t arising from direct mutations from the cell to S_{i+1},

- $B_i(t)$ to be the probability that a cell born in S_i at time 0 has no descendants in S_r at time t arising from mutations to S_{i+1} from descendants of the cell in S_i.

We should note that the probability that a cell born in S_i at time 0 has no descendants in S_r at time t equals $C_i(t)B_i(t)$.

We condition on the first birth to the cell in S_i at time s. There are then two cells. The probability that the new cell has no offspring equals $B_i(t-s)C_i(t-s)$. Any descendants of this cell are descendants of the birth process from the original cell. However, only descendants from future births to the original cell count as descendants of the birth process from the original cell. We thus have

$$B_i(t) = \int_0^t B_i(t-s)C_i(t-s)B_i(t-s)\mu_i e^{-\mu_i s}\,ds + e^{-\mu_i t}$$

$$= \mu_i e^{-\mu_i t}\int_0^t B_i^2(u)C_i(u)e^{\mu_i u}\,du + e^{-\mu_i t}.$$

Dropping the subscript i for the moment and multiplying both sides by $e^{\mu t}$ and differentiating, we obtain

$$[\mu B + \dot{B}]e^{\mu t} = \mu C B^2 e^{\mu t}.$$

Setting $D = 1/B$ and multiplying each side by $e^{-\mu t}$ gives

$$[\mu D - \dot{D}]e^{-\mu t} = \mu e^{-\mu t}C = \frac{d}{dt}[e^{-\mu t}D] \quad \implies \quad -e^{-\mu t}D = \int_0^t \mu e^{-\mu s}C\,ds - 1, \quad (2)$$

since $D(0) = 1/B(0) = 1$. Conditioning on the first mutation from the cell to S_{i+1} we obtain

$$C_i(t) = \int_0^t B_{i+1}(t-s)C_{i+1}(t-s)C_i(t-s)\lambda_i e^{-\lambda_i s}\,ds + e^{-\lambda_i t}$$

$$= \lambda_i e^{-\lambda_i t}\int_0^t B_{i+1}(u)C_{i+1}(u)C_i(u)e^{\lambda_i u}\,du + e^{-\lambda_i t}.$$

Multiplying both sides by $e^{\lambda_i t}$ and differentiating we obtain

$$e^{\lambda_i t}[\lambda_i C_i + \dot{C}_i] = \lambda_i e^{\lambda_i t}B_{i+1}C_{i+1}C_i \quad \implies \quad \frac{d}{dt}\ln C_i = [\lambda_i B_{i+1}C_{i+1} - \lambda_i]. \quad (3)$$

Solving equations (2) and (3) gives the following result.

Theorem 2. *It holds that*

$$B_i(t) = \frac{e^{-\mu_i t}}{1 - \int_0^t \mu_i e^{-\mu_i s}C_i(s)\,ds},$$

$$C_i(t) = e^{-\lambda_i t}\exp\left(\lambda_i \int_0^t B_{i+1}(s)C_{i+1}(s)\,ds\right).$$

References

ARMITAGE, P. AND DOLL, R. (1957). A two-stage theory of carcinogenesis in relation to the age distribution of human cancer. *British J. Cancer* **11,** 161–169.

CONNOLLY, R. B. AND KIMBELL, J. S. (1994). Simulation of cell growth governed by stochastic processes: application to clonal growth cancer models. *Toxicology Appl. Pharmacology* **124,** 284–295.

KINGMAN, J. F. C. (1975). The first birth problem for an age-dependent branching process. *Ann. Prob.* **3,** 790–801.

PORTIER, C. J., KOPP-SCHNEIDER, A. AND SHERMAN, C. D. (1996). Calculating tumor incidence rates in stochastic models. *Math. Biosci.* **135,** 129–146.

PORTIER, C. J., SHERMAN, C. D. AND KOPP-SCHNEIDER, A. (2000). Multistage stochastic models of the cancer process: a general theory for calculating tumor incidence. *Stoch. Environ. Res. Risk Assess.* **14,** 173–179.

ZHENG, Q. (2008). Stochastic multistage cancer models: a fresh look at an old approach. In *Handbook of Cancer Models with Applications*, World Scientific, Singapore, pp. 25–44.